# Empirical Study of Drive-by-Download Spyware

**Mark Barwinski, Cynthia Irvine and Tim Levin**
**Naval Postgraduate School, Monterey, USA**
markbarwinski@hotmail.com;
irvine@nps.edu;
levin@nps.edu;

**Abstract:** The ability of spyware to circumvent common security practices, surreptitiously exporting confidential information to remote parties and illicitly consuming system resources, is a rising security concern in government, corporate, and home computing environments. While it is the common perception that spyware infection is the result of high risk Internet surfing behavior, our research shows main-stream web sites listed in popular search engines contribute to spyware infection irrespective of patch levels and despite "safe" Internet surfing practices.
Experiments conducted in July of 2005 revealed the presence of spyware in several main-stream Internet sectors as evidenced in the considerable infection of both patched and unpatched Windows XP test beds. Although the experiment emulated conservative web surfing practices by not interacting with web page links, images, or banner advertisements, spyware infection of Internet Explorer based test beds occurred swiftly through cross-domain scripting and ActiveX exploits. As many as 71 different spyware programs were identified among 6 Internet sectors. Real estate and online travel-related web sites infected the test beds with, as many as 14 different spyware programs and one bank-related web site appeared to be the source of a resource consuming dialing program.
Empirical analysis suggests that spyware infection via drive-by-download attacks has thus far been unabated by security patches or even prudent web surfing behavior. At least for the moment, it appears the choice of web browser applications is the single most effective measure in preventing spyware infection via drive-by-downloads.

**Keywords:** Spyware, drive-by-download, malware, infection, internet, information assurance.

## 1. Introduction

Internet-based cyber attacks have been increasing in both frequency and complexity, and a strong emphasis on wealth appropriation through illegal means has taken over the once ideals- or publicity-driven hacker activities of the 1980's and 1990's. A spyware industry is well established and flourishes where legal and ethical issues are gray. It has become one of the greatest threats to cyberspace at a time of increased reliance on internetworking.

The predominant attack vector used by spyware today is through users' vulnerable web browsers. Insidiously, *drive-by-download* attacks require no action by the user other than to simply view a malicious or undermined web site. A prime example of this occurred in 2004, when the "Download.ject attack" compromised the web sites of numerous banks, insurance companies, auction outlets, and other main stream businesses (Krebs 2004). Visitors to these sites became infected by the mere action of going to the site. The attack installed key logging and Trojan horse software in visitors' computers and captured sensitive information such as Social Security Numbers, credit card numbers, user names, passwords, and encrypted financial communications (CNN 2004, Register 2004, Microsoft 2004).

We present an empirical analysis of drive-by-download attacks which shows the presence of spyware in several "low-risk" Internet sectors, including banking, online travel, and real estate. We also describe the variability of spyware susceptibility based on security patch maintenance practices and the type of browser used.

### 1.1 Motivation

Common wisdom dictates that high-risk behavior on the Internet leads to infection by spyware, viruses, Trojan horses, key loggers and the like. The use of peer-to-peer file sharing networks, the downloading of freeware and shareware, and the visiting of hacker- or warez-related web sites, as well as adult entertainment and gambling-related web sites might be considered "high risk behavior". However, there is evidence of risk in connection to mainstream web sites, despite the general perception that they are "safe" or "low risk" activities.

## 2. Background

With the development of HTML and the explosion of the global network communications infrastructure, web browsers became the dominant applications for exploring the Internet. As browsers battled for market share,

# Report Documentation Page

*Form Approved*
*OMB No. 0704-0188*



| 1. REPORT DATE | 2. REPORT TYPE | 3. DATES COVERED |
|---|---|---|
| **MAR 2006** | | **00-00-2006 to 00-00-2006** |

| 4. TITLE AND SUBTITLE | 5a. CONTRACT NUMBER |
|---|---|
| **Empirical Study of Drive-by-Download Spyware** | 5b. GRANT NUMBER |
| | 5c. PROGRAM ELEMENT NUMBER |
| 6. AUTHOR(S) | 5d. PROJECT NUMBER |
| | 5e. TASK NUMBER |
| | 5f. WORK UNIT NUMBER |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Postgraduate School ,Center for Information Systems Security Studies & Research (CISR),Monterey,CA,93943** | 8. PERFORMING ORGANIZATION REPORT NUMBER |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) | 10. SPONSOR/MONITOR'S ACRONYM(S) |
| | 11. SPONSOR/MONITOR'S REPORT NUMBER(S) |

12. DISTRIBUTION/AVAILABILITY STATEMENT
**Approved for public release; distribution unlimited**

13. SUPPLEMENTARY NOTES
**Proc. International Conference on i- Warfare and Security, Eastern Shore MD, 15-16 March 2006 pp.1-12**

14. ABSTRACT
**The ability of spyware to circumvent common security practices, surreptitiously exporting confidential information to remote parties and illicitly consuming system resources, is a rising security concern in government, corporate, and home computing environments. While it is the common perception that spyware infection is the result of high risk Internet surfing behavior, our research shows main-stream web sites listed in popular search engines contribute to spyware infection irrespective of patch levels and despite ?safe? Internet surfing practices. Experiments conducted in July of 2005 revealed the presence of spyware in several main-stream Internet sectors as evidenced in the considerable infection of both patched and unpatched Windows XP test beds. Although the experiment emulated conservative web surfing practices by not interacting with web page links, images, or banner advertisements, spyware infection of Internet Explorer based test beds occurred swiftly through cross-domain scripting and ActiveX exploits. As many as 71 different spyware programs were identified among 6 Internet sectors. Real estate and online travel-related web sites infected the test beds with, as many as 14 different spyware programs and one bank-related web site appeared to be the source of a resource consuming dialing program. Empirical analysis suggests that spyware infection via drive-by-download attacks has thus far been unabated by security patches or even prudent web surfing behavior. At least for the moment, it appears the choice of web browser applications is the single most effective measure in preventing spyware infection via drive-by-downloads.**

15. SUBJECT TERMS

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT **unclassified** | b. ABSTRACT **unclassified** | c. THIS PAGE **unclassified** | **Same as Report (SAR)** | **12** | |

new features were added, such as support for JavaScript and Java applets in 1995, Cascading Style Sheets (CSS) in 1996, and the Document Object Model (DOM) in 1997. Together these technologies led to the realization of dynamic content.

Concurrently with the growth of the Internet, the number of reported software vulnerabilities and computer incidents has grown exponentially. Figure 1 depicts trends in both vulnerabilities and incidents between 1988 and 2004, soon after the introduction of dynamic web content.

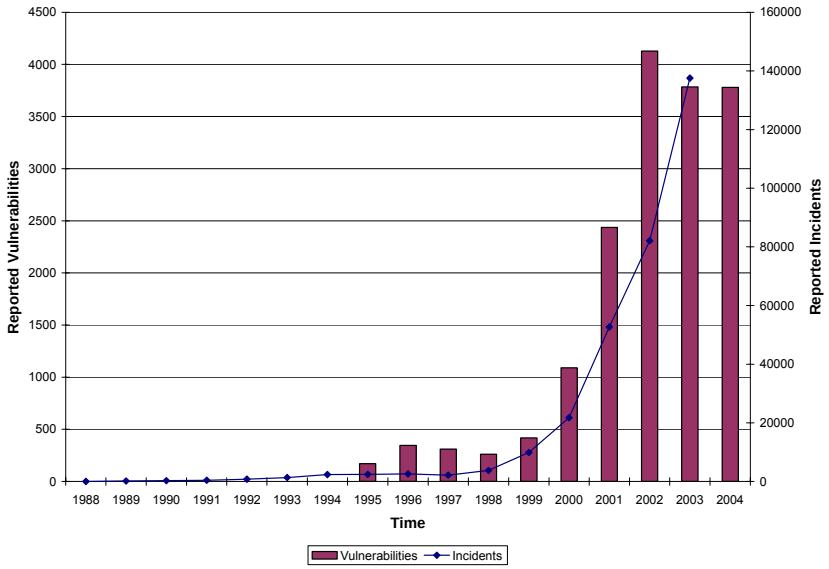


**Figure 1:** Rise of reported vulnerabilities and incidents (CERT).

The term "spyware" was first used in an October 1995 newsgroup forum to partially describe Microsoft Corporation's business model. Soon thereafter, subsequent postings in various newsgroups started using the term to describe malicious software not fitting squarely within the definition of a virus program. Steve Gibson wrote the first anti-spyware program in 2000, soon after the arrival of "Elf Bowling" in 1999, a popular game bundled with tracking software. Since then, spyware has exploded into a multi-million dollar industry.

## 3. Spyware definition

A standard definition of the term "spyware" has proven elusive, and its meaning has varied greatly. At one end of the spectrum, spyware is limited to the collection of personally identifiable information, such as key logging, and password stealing. At the other end of the spectrum, spyware has been defined as software collecting practically any information from a system, and forwarding it to a third party in a manner unknown to the computer user. Unfortunately, the latter definition would encompass such programs as Microsoft AutoUpdate and anti-virus updating utilities, programs with a clear benefit to the user.
Attempts to better define spyware have also led to confusing terms such as snoopware, scumware, junkware, thiefware, parasite software, undesirable software, and others.

### 3.1 Convergence of activities

We distinguish spyware by the convergence of a common set of behaviors or *activities* in a software program deployed to profit financially or strategically from data gathering activities. These activities consist of the ability to operate in the background, collect information, communicate this information to a third party, and maintain a presence in a computer system. In short: hide, collect, communicate, and survive in a hostile environment.

#### *3.1.1 Hide*

Spyware software must be able to hide, at least in part, the mechanisms associated with its installation, execution, data collection, or communication. In legitimate programs, "hiding" can be seen as the desirable aspect of staying out of the user's way. But in spyware, program installation or process hiding is accomplished via the exploitation of various system vulnerabilities. Spyware also utilizes obfuscating naming conventions. Hiding among legitimate system files, spyware uses file names similar to those used by software vendors. These files are stored in folders associated with legitimate software products. System,

font, and temporary folders, to name a few, offer the added benefit of containing sometimes thousands of files, thus providing ample hiding opportunities. Additionally, collected data may be hidden in encrypted files, in the system registry, or in unallocated sectors of the hard drive.

Finally, spyware communications may be hidden by encrypting the transmission, by performing sparse, limited transmissions, or by hi-jacking the transmission medium of an application, which possesses legitimate Internet access.

### *3.1.2 Collect*

Spyware must also be able to collect information from the infected host. This information may range from relatively benign non-identifiable market demographics, to sensitive personally identifiable demographics, to highly valuable and desirable targeted information such as financial or medical information, corporate trade secrets, or sensitive or classified government information.

### *3.1.3 Communicate*

Communication between spyware on the infected host and, one or more remote collection points, is essential. Communication channels may include high speed wired and wireless network connections such as Ethernet, wireless 802.11, and Bluetooth technologies, as well as dial-up modem connections.

### *3.1.4 Survive*

The last basic activity of spyware is its ability to survive. Spyware resides in a hostile environment. Implicit in the desire to remain undetected is the consequence that when detected, users will act to remove it. There is a high likelihood that attempts will be made to remove or disable the spyware software at some time during the deployment and maintenance phase of its lifecycle. Therefore, spyware must be resilient, remaining within the compromised system for as long as possible, e.g., through redundant processes and multiple re-installation vectors in the system registry.

## 3.2 Spyware usage sectors

Three primary usage sectors are identified: *marketing*, *surveillance*, and *resource consumption*. The marketing area is defined as any business making use of demographic information, whether it is anonymous or directly identifies a particular user, strictly for the purpose of selling a legitimate product or service. The surveillance category is defined as having as its main objective the tracking of users or the gathering of user information in a far greater degree of detail than the marketing category. Its main use is in law enforcement, industry asset and employee monitoring, or intelligence gathering activities. Resource consumers are defined as those who financially benefit from utilizing system resources in compromised systems. Figure 2 depicts the distribution of spyware among the three main usage sectors.

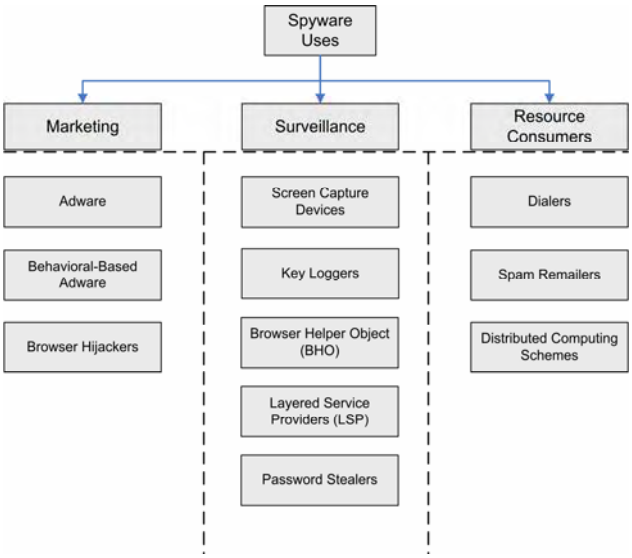


**Figure 2:** Spyware uses and mechanisms

Based on these definitions, adware and behavioral-based advertising (both of which may include pop-up advertising-type behavior), and browser hijackers fall within the marketing category. They draw traffic to affiliated web sites and attempt to generate business transactions.

In the surveillance category, screen capture devices, key loggers, password stealers and similar programs closely monitor user activities on a system. Browser Helper Objects and Layered Service Providers are able to intercept web traffic before Secure Sockets Layer (SSL), thus further expanding surveillance capabilities, encrypts it.

Resource consuming spyware may profit from distributed computing by taking system resources away from the user. For example, re-mailers utilize bandwidth-rich DSL-connected systems to generate and distribute spam. Another yet more egregious instance utilizes unused storage or CPU cycles in a compromised system and sells them to clients with massive processing or storage requirements.

## 4. Experiment methodology

The experiment was intended to assess generally "safe" sectors of the Internet for the potential of spyware infection as a result of drive-by-download attacks. The assessment of these Internet sectors was accomplished via passive; "safe" web surfing activities. Links associated with banking, insurance, children, real estate, online travel, universities, government, and military-related web sites were evaluated. Additionally, high-risk areas of the Internet including online gambling, hacker and "warez," and adult entertainment-related web sites, were also evaluated for comparison purposes. Safe web surfing activities consisted of limited interaction with the web site so as to avoid accepting, authorizing, or inviting installation of software. Banner advertisements, images, and links within these web sites were not clicked upon. Requests for acceptance of certificates, and download or execution of programs or browser plug-ins were dismissed. The population of mainstream web sites was compiled from search engine queries pertaining to a specific industry and from specific listings. The intent behind this experiment design was to replicate activities that might be conducted by an average prudent user who does not intentionally connect to what may be considered high risk areas of the Internet.

As it pertains to spyware, a *drive-by-download* is an attack conducted by a malicious web site in which spyware is installed in a victim's computer. This is accomplished without alerting, or requiring authorization or overt action from the user.
The empirical analysis of spyware consisted of the following activities:

- Compile approximately 500 web site links for each of eight safe and three unsafe sectors of the Internet.
- Collect system snapshot data prior to the commencement of surfing simulation in order to establish a baseline.
- Visit each web site simultaneously with four different virtual machines. The virtual machines consisted of patched and unpatched Windows XP operating systems with Internet Explorer and patched and unpatched Windows XP operating systems with Firefox.
- Collect system information following each visited link.
- Collect system snapshot upon conclusion of web surfing simulation in each Internet sector for later comparison against baselines.
- Identify malicious web sites responsible for infection.
- Conduct comparative analysis among patch levels, Internet browsers, and anti-spyware scanning tools.

The experiment was conducted using a collection of VBScripts. Scripts were used to collect system baselines prior to the commencement of web surfing activities. Additional scripts drove browsers to various Internet sector Uniform Resource Locators (URLs) where 15-seconds to download a web page and 5-second idle time allowed spyware infections to commence. The scripts collected system snapshots prior to visiting the next URL. These system snapshots were later compared against the baselines.

Infection detection was accomplished via three different techniques. The first consisted of the use of a host integrity monitoring system. Baseline snapshots collected by the host integrity monitoring system were compared against snapshots collected at the conclusion of the experiment. This provided information on changed files, folders, user accounts, running services, and open communication ports.

The second technique employed client-based anti-spyware scanning tools: Microsoft AntiSpyware (Beta 1), Lavasoft Ad-Aware, Spybot Search and Destroy, and Earthlink SpyAudit. These tools were used to determine spyware infection at the completion of each Internet sector experiment.

The third technique used a set of client-based third-party tools used to collect system information after visiting each individual web site. The collected information included a list of running processes and services, open communication ports and files associated with such ports, a list of applications or programs scheduled to auto-start upon boot-up or login, browser security and preference settings, a snapshot of the hosts file, a list of browser favorites or bookmarks, and over 87 different system registry sub keys. This information was compared to baseline snapshots in an effort to identify specific changes caused by a particular web site.

The experiment collected data on the relative risk factor of various Internet sectors, the use of different browsers, the detection performance of the various anti-spyware tools, and the state of a default configuration unpatched Windows XP system versus a default configuration fully patched Windows XP system.

## 4.1 URL determination

Approximately 5,000 different URLs covering eleven different Internet sectors were collected for this experiment. A list of banking institution-related web sites was obtained from the Federal Deposit Insurance Corporation (FDIC). A list of child-related web sites was obtained in part from the American Library Association and their Great Web Sites Seal of Approval Program. Additional links were obtained at other minor child-related directories. University-related links were compiled from the University of Texas at Austin, which maintains an alphabetical list of all U.S. community colleges, and universities. Government and military-related web sites were compiled using the Google™ search engine. Searches were conducted by filtering for the .gov or .mil domains. Government-related web sites are defined to be web sites hosted by a federal or state government agency. A military-related web site is defined to be a web site hosted by a military-related agency or branch, in the .mil domain.

Web sites for the remaining Internet sectors were compiled by conducting key word-specific searches. For example, when compiling the online gambling sector of the Internet, the search query consisted of keywords “online gambling” or “online casino.” Real estate-related web sites were identified using keywords such as “real estate,” “realtor,” and “mortgage.”

Web sites compiled from search engine queries were stripped of their long URLs and restricted to their domains, allowing the test bed to visit a greater number of different web domains as opposed to several web pages within the same domain.

## 4.2 Unrelated URLs

The web-site selection methodology used for the experiment resulted in some URLs that were unrelated to their intended sectors. The criteria used in determining unrelated URLs consisted of:

- Web sites that do not sell or provide services or products associated with the industry in question.
- Web sites consisting of generic non-sector-specific content.
- Web sites that returned an invalid URL or request-error page within the requested domain.

An analysis of the number of unrelated URLs was performed by visually inspecting a representative sample, 75 URLs, from each of the insurance, child, real estate, and online travel-related sectors. Each of these sectors had a population of 500 web sites. Table 1 shows the number of false positives found per sector. Based on the sample, an estimate of the number of false positive URLs present in each sector is also provided. Calculations were made using the hypergeometric distribution model with at least a 95% confidence.

**Table 1:** False positive URLs by sector

| Sector | Sector Size | False Positives | False Positive URLs Estimate (for 500 web sites) | Confidence |
|---|---|---|---|---|
| Insurance | 75 | 1 | 1 to 30 | 0.95181 |
| Children | 75 | 9 | 30 to 101 | 0.95007 |
| Real Estate | 75 | 31 | 156 to 260 | 0.95029 |
| Online Travel | 75 | 2 | 3 to 40 | 0.95181 |

## 4.3 Test bed description

The test bed was comprised of a workstation, a file server, a hub, and a router. The workstation was configured with a fully patched Windows XP operating system, which hosted the integrity monitoring system and VMWare. The VMWare environment within this workstation was configured with five clients. The workstation consisted of an Intel Pentium 4 3.2Ghz, with 2GB of RAM and two 120GB hard drives. A Windows XP file server located on a separate computer was used for the storage of test data. This same server was also used for the collection of network traffic during the web surfing simulation phase of the experiment.

The virtual machines were equipped with common third party applications associated with the enhancement of the web surfing experience, such as Macromedia Shockwave, Macromedia Flash Player, and a Java runtime environment.

A router implementing Network Address Translation was used to protect the test bed from infection by means other than strictly spyware-related drive-by-download attacks. The VMWare system hosted five simultaneous virtual machine test platforms, each consisting of a separate Windows XP operating system. The virtual machines consisted of a passive experimental control identified as hostname PASSIVE, a default unpatched Windows XP and Internet Explorer installation identified as IE, a default unpatched Windows XP and Firefox installation identified as FF, a fully patched Windows XP and Internet Explorer installation identified as IESEC, and a fully patched Windows XP and Firefox installation identified as FFSEC (See Figure 3).

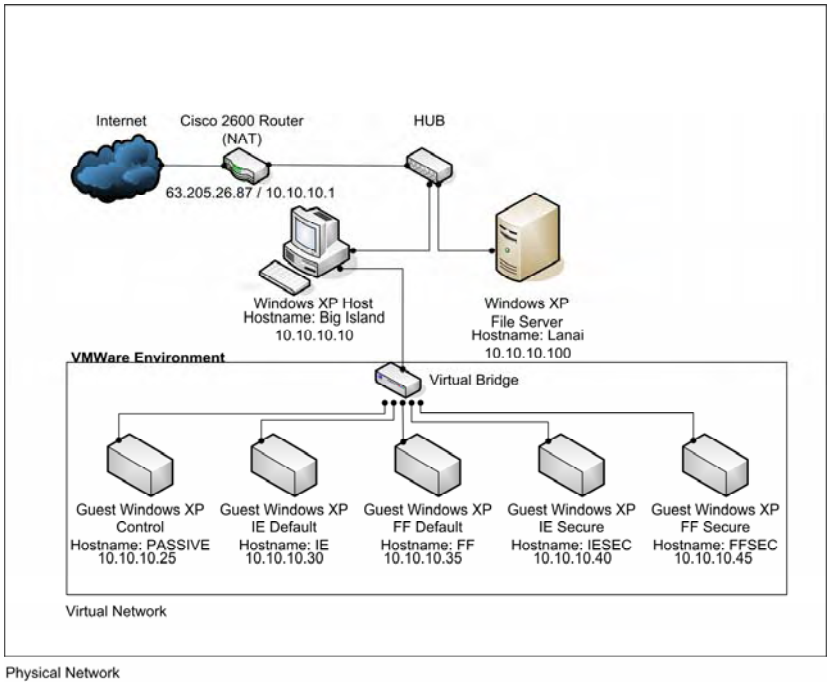


**Figure 3:** Test bed network topology

## 4.4 Infection validation

Preliminary tests were performed to determine if infection of the test bed using the proposed methodology was, in fact, possible, and to help determine various parameters for the experiment. The IE platform was directed to five known malicious web sites. Table 2 shows the infection download times as well as the time required to collect the system snapshot data following infection. This data and other measurements resulted in establishing a 20-second “visitation” parameter (see below).

Three web sites were found to take between 9.5 and 17.6 seconds to download, well within the combined 20-second window of time. Snapshot collection times for these three web sites ranged between 17.1 and 38.1 seconds, reflecting performance degradation from the ongoing spyware infection. The two remaining web sites utilized exploits that crashed the browsers within 10 seconds of arrival, halting benchmark times. All five web sites successfully infected the test bed with spyware.

**Table 2:** Preliminary malicious web site download comparisons

| URL | Download (seconds) | Collection (seconds) | Notes |
|---|---|---|---|
| www.unix-time-format.dzwonki.pruszkow.pl | 15.1 | 17.1 | Drive-By-Download |
| Viking-supply-net.to.opole.pl | N/A | N/A | Browser crashed |
| Food-pyramid.ok.opole.pl | N/A | N/A | Browser crashed |
| Sex-archive.biz/movies/ | 9.5 | 38.1 | Drive-By-Download |
| m.cpa4.org/reality | 17.6 | 21.6 | Drive-By-Download |

## 5. Analysis

The experiment demonstrated that out-of-the-box installations of Windows XP and Internet Explorer (IE test bed) were most susceptible to spyware infection. Figure 4 shows the breakdown of infections for the IE test bed among the various Internet sectors. The Hacker and Warez sectors led with the most spyware infections followed by the online travel and real estate Internet sectors. The adult entertainment sector followed close behind. Minor hits were noted for the banking and online gambling sectors and appeared to be associated with a single web site in each case. It is interesting to note that very similar results were recorded for the fully patched installation of Windows XP and Internet Explorer (IESEC test bed). Figure 5 shows the IESEC test bed had considerable infections for the adult entertainment, hacker and warez, and online travel sectors. Additionally, these two figures also show detection results among the four scanning tools used in this experiment. Reported infections increase from left to right starting with Earthlink SpyAudit, Ad-Aware, Spybot Search and Destroy, and Microsoft AntiSpyware, respectively. Both the IE and IESEC platforms were considerably infected not only by the high-risk sectors, but also by web sites found within the online travel sector. Three web sites were identified as malicious and responsible for the infections in this sector. Based on the name of the URLs, these web sites did not appear to be false positives.

Figure 6 shows a comparison of platform infection rates in each of the sectors. The Firefox based platforms did not experience spyware infection. Limited infection rates were noted for the IE platform in the banking, real estate, and insurance Internet sectors. Interestingly, the online travel sector had a greater number of spyware infections than the adult-entertainment sector.

Figure 7 groups the various Internet sectors by test bed, clearly showing the passive experiment control test bed and the Firefox test beds were not infected by spyware. Additionally, with the exception of the hacker and warez related web sites, combined infection counts for all four anti-spyware scanning tools appears to range between the low to high 30's. It is apparent from Figure 6 and Figure 7 that spyware infection by the adult entertainment, hacker/warez, and online travel-related sectors were not significantly diminished with the installation of Service Pack 2 and subsequent security patches.

Network traffic analysis revealed that many of the attacks consisted of JScripts invoking ActiveX objects and cross-domain vulnerability exploitation. Further analysis of the system snapshots revealed numerous common binaries encountered among various web sites and different Internet sectors. Table 3 provides a list of the most common malicious binaries identified during the course of the experiment and the servers from which they were downloaded. Many of these binaries are associated with such spyware programs as 180Search Assistant, Bargain Buddy, CoolWebSearch, ShopAtHome, MediaGateway and the like.

During the course of the experiment, a total of 16 malicious web sites were identified from among the eight different "safe" Internet sectors. Three malicious web sites were identified in the online travel Internet sector, 12 web sites in the real estate sector, and 1 web site in the bank Internet sector, respectively. Of these 16 web sites, 12 appear to be registered under the Polish Internet domain.

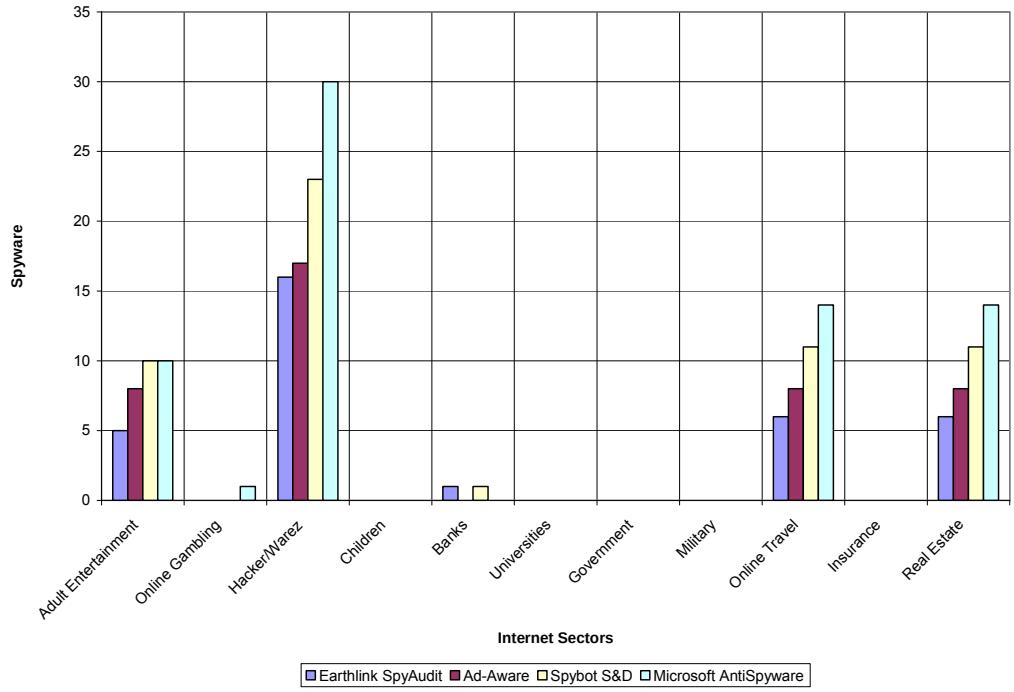


**Figure 4:** IE test bed spyware infection by sector

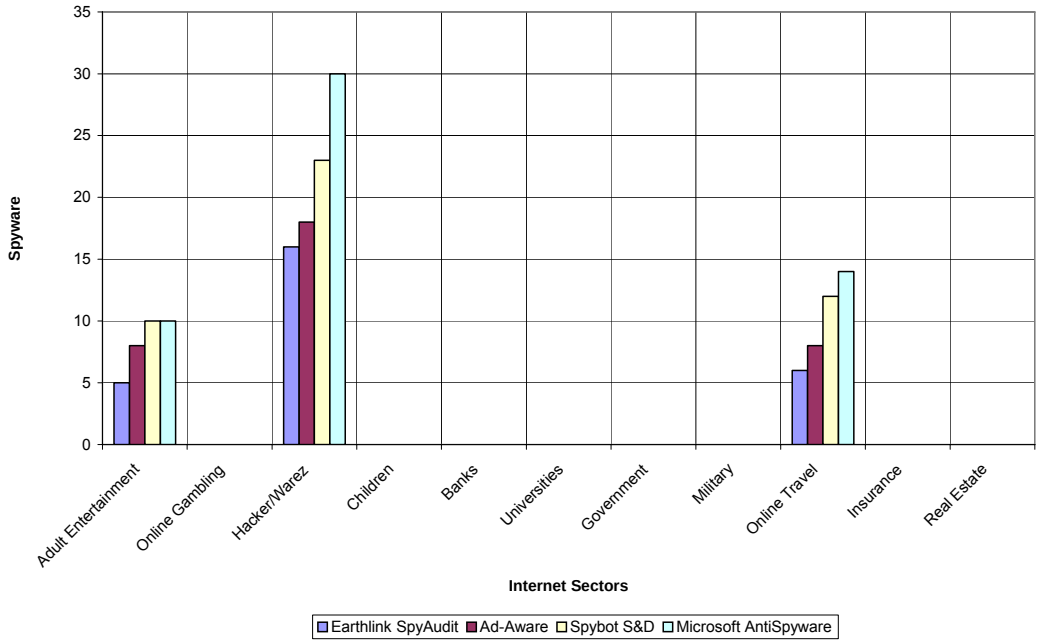


**Figure 5:** IESEC platform spyware infection by sector

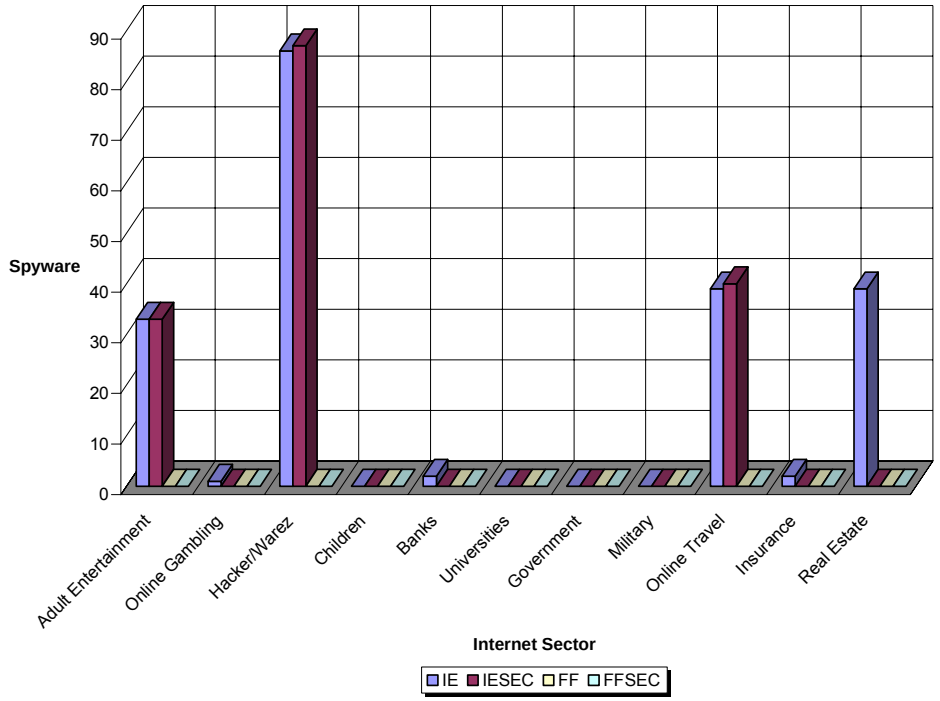


**Figure 6:** Test bed infection comparison by sector

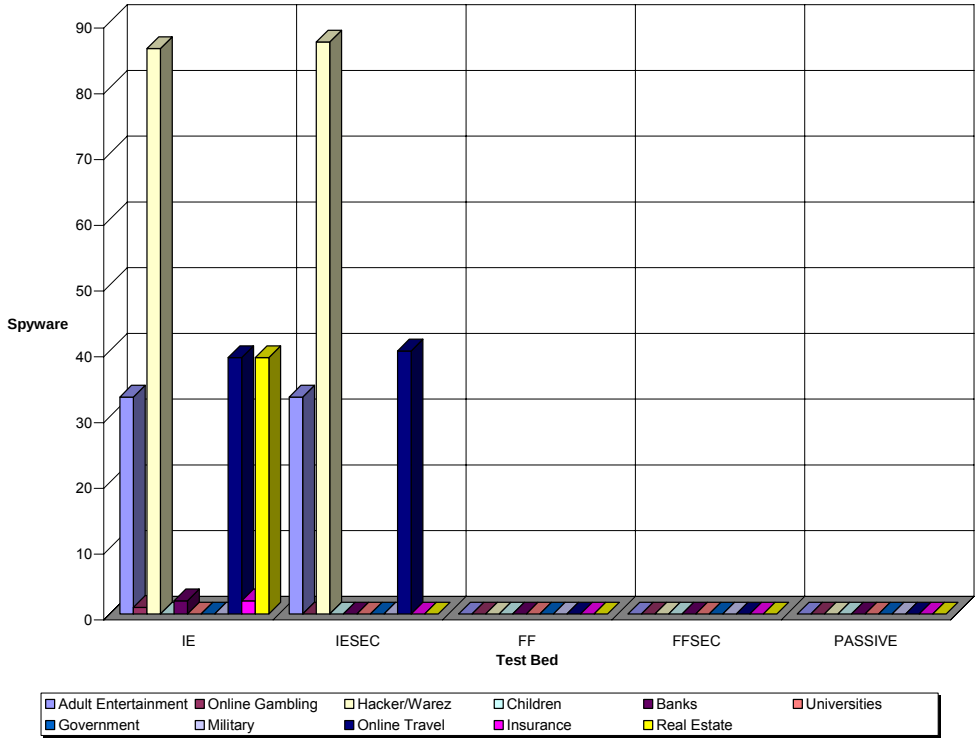


**Figure 7:** Test bed infection comparison by platform

**Table 3:** Observed infectious binaries and associated servers

| Malicious Binaries | Servers |
|---|---|
| Bundle_cdt1006.exe | tatic.flingstone.com |
| cdt1006.sah<br>setup4030.cab | Downloads.shopathomeselect.com |
| Toolbar3.cab | download.websearch.com |
| Bridge-c139.cab<br>bridge-c420.cab<br>MediaGateway.exe | static.windupdates.com |
| clienthook.dll | installs.180solutions.com |
| s.dll | gr2.cc |
| Nem220.dll | cdn2.movies-etc.com |
| agentprefs2.sah<br>validate.sah<br>global.sah | www.shopathomeselect.com |
| optimize.exe | cdn.climaxbucks.com |
| ProSiteFinder.exe | sds.qckads.com |

## 6. Related work

The University of Washington Computer Science and Engineering department conducted a study on the presence of spyware within their network of 40,000 to 50,000 hosts (Saroiu 2004). This study showed that spyware had successfully penetrated most organizational boundaries; 69% of the organizations within the university environment contained at least one host infected with at least one variety of spyware.

Microsoft Corporation has also undertaken research in this area (Wang 2004, Wang 2005), in which the concept of Auto-Start Extensibility Points (ASEPs) is introduced and a framework is established for the detection of spyware infection via the monitoring of these ASEPs.

Wang et al. manipulated Internet Explorer programmatically to visit malicious web sites (Wang 2005). Approximately 5000 potentially malicious URLs were visited. Through the use of virtual machines, various patch levels were used, starting with the lowest patch level (Windows XP SP1 unpatched). Upon infection, progressively higher patch level virtual machines were tested until a malicious web site was documented to have exploited fully patched systems, leading Microsoft to conclude that a zero-day exploit had been encountered.

In 2004, researchers conducted an experiment in which a total of 600 web sites were visited (Shukla 2005). Testers interacted with each site in a manner to "simulate the behavior of naïve users." The experiment was organized into four sectors of the Internet – E-Commerce, Recreation and Entertainment, Download Search and Directory, and News and Education related web sites. The study used Ad-Aware and Spybot Search and Destroy to detect spyware infection. The study concluded that user browsing behavior is "responsible for much of the spyware dissemination on computers."

## 7. Conclusions

Spyware has penetrated personal, business and government systems despite common defense-in-depth approaches. Up-to-date patched computer systems, firewalls, and anti-virus programs have thus far failed to stem the tide of spyware infection. Empirical analysis shows spyware infection is possible by visiting "safe" sectors of the Internet (e.g., banking, online travel, and real estate-related web sites) while practicing "safe" passive web surfing activities. Infection by "high risk" sectors of the Internet was also confirmed during the course of the experiment.

### 7.1 Internet sectors

Child-related, banking, university, government, military, online travel, insurance, and real estate-related web sites were evaluated for the presence of drive-by-download spyware.

Consistent with commonly given advice, adult entertainment, and hacker and warez-related web sites were found to make use of drive-by-download techniques and infected both patched and unpatched versions of Internet Explorer, with the latter sector showing the greatest number of infections.

Both the online travel and the real estate sectors of the Internet showed spyware infection greater than that observed in the adult entertainment Internet sector. Lastly, a single infection was also noted in the banking sector.

### 7.2 Browser performance

Comparison of browsers with respect to the likelihood of infection by spyware through drive-by-downloads revealed infection to be strictly limited to Microsoft Internet Explorer. It is suspected that many spyware programs use ActiveX exploits to gain access to victim computers. Firefox does not natively support ActiveX. Additionally, it is suspected greater emphasis is placed on the development of spyware for an Internet browser, which holds approximately 90% market share on the Internet. Firefox currently holds approximately 8% of the browser market. Finally, out-of-the-box default security settings may not be equivalent for the two browsers, contributing to the dramatically different results.

### 7.3 Patch performance

Empirical analysis of the various test platforms showed that operating system and browser patching made little difference in the spyware infection rates. Comparisons between a default installation of Windows XP and a fully-patched installation which included Service Pack 2 and numerous other operating system and browser patches revealed similar infection behavior.

A plausible explanation for these findings may reside in the use of overly permissive browser security configuration settings. The experiment used browser settings typical of those commonly used or with attractive functionality, such as allowing ActiveX and other script execution, Java applets, Flash and Shockwave. In this manner, platform configurations were representative of commonly found systems in the real world.

### 7.4 Anti-spyware scanning tools

Four different and freely available scanning tools were used to detect spyware infection. Empirical analysis of the virtual machines and comparative analysis among the four scanning tools shows that Microsoft's AntiSpyware consistently reported a higher number of infections, followed by Spybot Search and Destroy, Ad-Aware, and SpyAudit. Since there is no standard way of reporting results among scanning tools, these results are not necessarily indicative of the detection rate for a given tool but instead may simply highlight reporting differences among vendors.

## References

Barwinski, M.A. (2005) Taxonomy Of Spyware And Empirical Study Of Network Drive-By-Downloads, Master's Thesis, Naval Postgraduate School, Monterey, CA, USA, September 2005.

Krebs, Brian (2004). PC Users Warned of Infected Web Sites, Washington Post. Available at www.washintonpost.com/wp-dyn/articles/A5524-2004Jun25.html (Last accessed July 11, 2005).

CNN (2004). Trojan virus attacks popular web sites. Available at http://www.cnn.com/2004/TECH/internet/06/25/internet.attack/ (Last accessed September 12, 2005)

Register (2004). Bofra Exploit Hits Our Ad Server Supplier. Available at http://www.theregister.co.uk/2004/11/21/register_adserver_attack/ print.html (Last accessed September 12, 2005)

Microsoft (2004). What You Should Known About Download.Ject. Microsoft Corporation. Available at http://www.microsoft.com/security/incident/download_ject.mspx, June 2004. (Last accessed September 12, 2005)

Saroiu, Stefan; Gribble, Steven D.; Levy, Henry M. (2004). Measurement and Analysis of Spyware in a University Environment, in Department of Computer Science & Engineering, University of Washington. (Proceedings of the 1st Symposium on Operating Systems Design and Implementation (NSDI), San Francisco, CA March 2004) Available at http://www.cs.toronto.edu/~stefan/publications/ nsdi/2004/spyware.html. (Last accessed March 14, 2005).

Shukla, Sudhindra, Fui-Hoon Nah, Fiona (2005). Web Browsing and Spyware Intrusion. Communications of the ACM. Vol. 48. No. 8. pp 85.

Wang , Yi-Min, Beck, Doug, Jiang, Xuxian, Roussev, Roussi (2005). Automated Web Patrol with Strider HoneyMonkeys: Finding Web Sites That Exploit Browser Vulnerabilities. Technical Report MSR-TR-2005-72. Cybersecurity and Systems Management Research Group, Microsoft Research, Redmond, Washington. June 4, 2005.

Wang, Y., Roussev, R., Verbowski, C., Johnson, A., Wu, M., Huang, Y., Kuo, S. (2004). Gatekeeper: Monitoring Auto-Start Extensibility Points (ASEPS) for Spyware Management, USENIX Association Proc XVIII LISA 2004.